

TRABAJO FINAL DE GRADO EN TRADUCCIÓN E INTERPRETACIÓN

TREBALL FINAL DE GRAU EN TRADUCCIÓ I INTERPRETACIÓ

Departament de Traducció i Comunicació

TÍTULO / TÍTOL

**Análisis del español original y traducido en los
discursos del Parlamento Europeo y el Congreso de los
Diputados**

Autor/a: M.^a Elvira Lacruz Puchades

Tutor/a: María Calzada Pérez

Universitat Jaume I, Castellón

Fecha de lectura/Data de lectura: junio 2015



Resumen/ Resum:

En este trabajo de fin de grado se examina el español de los corpus del Congreso de los Diputados (CD) y del Parlamento Europeo (PE) procedente del archivo del grupo de investigación ECPC. El español del CD es, lógicamente, original, mientras que el del PE puede ser tanto original como traducido. El análisis del presente trabajo se aborda desde un punto de vista tanto conceptual como gramatical. Este último ha sido posible gracias a una aportación que he realizado al corpus, el etiquetado morfológico de todas las palabras, llamados POS (*Part of Speech*). Asimismo compararé el lenguaje puramente institucional y parlamentario de los corpus elegidos con el lenguaje común que se utiliza día a día, para lo cual me he valido del listado de las 10 000 palabras más frecuentes de nuestro idioma, según la RAE. Este trabajo, por tanto, se propone, de un lado, confrontar el español original con el traducido en este tipo de contextos y, de otro, conocer más de cerca cómo se trabaja con los corpus, qué información nos pueden aportar estos, para qué son útiles y sobre todo cuál es su aplicación en el campo de la traducción y la interpretación.

Palabras clave/Paraules clau: (5)

Corpus	Parlamento Europeo
Español original	Congreso de los Diputados
Español traducido	

Índice

Análisis del español original y traducido en los discursos del Parlamento Europeo y el Congreso de los Diputados	3
Motivación y justificación	3
Contextualización	3
Metodología.....	6
Variación léxica	9
Listas de frecuencia	10
Palabras clave	11
Concordancias.....	12
Etiquetado morfológico	13
Resultados.....	13
Variación léxica	13
Listas de frecuencia	14
Parlamento Europeo..	14
Congreso de los Diputados.....	15
Palabras clave	16
Etiquetado morfológico (<i>POS</i>)	20
Conclusiones.....	22
Bibliografía.....	24

Análisis del español original y traducido en los discursos del Parlamento Europeo y el Congreso de los Diputados

Motivación y justificación

La realización de este trabajo no fue una elección propia y premeditada; fue una propuesta que me planteó mi tutora. Decidí aceptar el reto porque me interesaba explorar un tema desconocido para mí: la lingüística de corpus (LC) y sus procedimientos metodológicos. Aunque, en los primeros años de carrera, tratamos los corpus de forma general, nunca había trabajado con ellos tan profundamente como en esta ocasión. En consecuencia, gracias a esta asignatura he podido aprender o refrescar, por un lado, usos generales de los corpus así como las herramientas de etiquetado y concordancias y, por otra parte, he descubierto aspectos del español original y traducido en contextos parlamentarios básicamente monolingües (el Congreso de los Diputados, CD) o multilingües (el Parlamento Europeo, PE).

Contextualización

Para contextualizar este trabajo debemos hablar por una parte de la lingüística de corpus (LC) y, por otra, de las dos grandes instituciones ante las que se pronuncian los debates de los textos que se examinan en este trabajo: el Parlamento Europeo y el Congreso de los Diputados.

Según Sinclair un corpus es “a collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language” (Sinclair, 1996, p.4, en Vargas, 2012, p.68).

Por su parte, para Vargas:

La lingüística de corpus es un campo de investigación caracterizado en la actualidad por dos aspectos básicos y fundamentales: estudiar de forma empírica la lengua en uso con el fin de describirla o analizar algún aspecto intra-/inter-textual o discursivo con un fin concreto; y utilizar el ordenador para el almacenamiento y el análisis de los datos. (Vargas, 2012, p.67)

Por tanto, la relación que existe entre traducción y lingüística de corpus es notable. Gracias a estas recopilaciones de textos, la traducción se ve positivamente reforzada ya que los corpus nos ayudan a ver cómo se comporta el lenguaje, qué expresiones se usan,

y en el caso de tratarse de un corpus paralelo, como veremos más adelante, nos presentará opciones de traducción.

Existen diversos tipos de corpus a los que Vargas (2012) se ha acercado de forma especialmente esclarecedora. Según esta autora, si tenemos en cuenta su contenido, se puede decir que un corpus es general o específico; según su tamaño, puede clasificarse como grande, mediano o pequeño; según la codificación del texto o el nivel de procesamiento interno, se distingue entre corpus en bruto (si el texto no contiene anotación de ningún tipo) o etiquetado (bien porque contiene etiquetas descriptivas de los elementos constitutivos de cada texto, bien porque las etiquetas explicitan aspectos lingüísticos de cada ítem del corpus, como podría ser su categoría gramatical). Si tenemos en cuenta las lenguas que se recopilan, un corpus puede ser monolingüe, bilingüe o multilingüe. Si estamos ante un corpus bilingüe o multilingüe, el corpus puede ser comparable o paralelo. En el primer caso, los textos que lo componen comparten determinados rasgos como por ejemplo, el tema, el tipo de texto, el periodo de tiempo en que se redactaron los textos, la función comunicativa, el grado de especialización, etc. Por otra parte, un corpus paralelo es aquel que contiene textos redactados en una lengua (original) y está acompañado de las traducciones de estos a otras lenguas.

En cuanto a las dos instituciones cuyos discursos se estudian aquí (Parlamento Europeo, PE, y Congreso de los Diputados, CD), cabe realizar también unos breves comentarios. El PE se enmarca dentro de la Unión Europea ya que constituye, junto al Consejo Europeo, el Consejo de la Unión Europea, la Comisión Europea, el Tribunal de Justicia, el Tribunal de Cuentas y el Banco Central Europeo, una de las principales instituciones de la UE.

Tal y como se recoge en la página web oficial del Parlamento Europeo, los diputados elegidos por sufragio universal participan en debates públicos y entre sus principales ámbitos de trabajo destacan la legislación y el presupuesto. El PE es el único órgano de la UE que se elige por sufragio directo y que supone una democracia activa con un doble carácter; multinacional, puesto que representa a los 500 millones de habitantes de la UE y multilingüe, porque tiene la obligación de garantizar el mayor nivel de multilingüismo. Cualquier ciudadano europeo puede presentarse a las elecciones al PE

y, por tanto, no se puede exigir a los diputados que tengan un dominio excelente de las lenguas más utilizadas, en este caso, inglés o francés. Todos los diputados del PE tienen el derecho a poder seguir los debates, escribir y expresarse en cualquiera de las lenguas oficiales así como a que se les traduzca. Actualmente en el PE se utilizan 24 lenguas oficiales, un gran reto lingüístico. Estas 24 lenguas oficiales permiten 552 combinaciones lingüísticas ya que cada lengua puede traducirse a las otras 24. (Parlamento Europeo, 2015)

Tal y como explica María Valdivieso Blanco en su artículo «Política Lingüística, norma lingüística y terminología en el plano supranacional. La unión europea», este multilingüismo tiene unas repercusiones que se ven plasmadas en la existencia de grandes servicios de traducción e interpretación, en el desarrollo de un sociolecto, la *eurojerga*, y en una serie de medidas encaminadas a garantizar la calidad de la redacción jurídica y la coherencia terminológica.

La *eurojerga* es un sociolecto que crea una lengua híbrida cargada de interferencias (a través de la traducción), que proceden de la principal lengua de origen, actualmente el inglés (Valdivieso, 2008). Este sociolecto contiene un léxico específico propio que se acuña en la UE para designar los términos, neologismos concretos, siglas etc. que surgen en este contexto europeo. Lo que ocurre con estos términos es que si los sacamos de contexto o si no tenemos conocimiento suficiente y profundo sobre la comunidad europea son difíciles de entender. Por esta razón, la *eurojerga* ha recibido críticas (Koskinen, 2000).

Al mismo tiempo, los textos europeos muestran una tendencia a la simplificación léxica, morfológica y sintáctica con el fin de obtener cierta uniformidad entre las lenguas (Valdivieso, 2008). Esta tendencia la observamos en los internacionalismos (palabras que muestran un parecido formal en distintos idiomas) como ocurre con la palabra *corrupción* (*corruption* en inglés y francés), que se ha traducido así en vez de por *cohecho* (un término más oportuno).

La falta de claridad de algunos vocablos y el estilo que define los documentos europeos es lo que ha ocasionado numerosas críticas sobre el lenguaje de las instituciones comunitarias en general. No obstante, tal y como plantea María Valdivieso en su artículo, se podría establecer una dicotomía en esta crítica. Por una parte, nos

llevaría a bastante confusión el empleo de términos más generales, universales y conocidos por todos para explicar situaciones y realidades propias del contexto europeo. Por otra parte, se debe reconocer que muchas veces se emplean unos términos poco claros que no vienen justificados por ninguna realidad concreta (Valdivieso, 2008).

La “contrapartida” nacional del PE es el Congreso de los Diputados:

El Parlamento español, que recibe el nombre de Cortes Generales, es bicameral. Está formado por el Congreso de los Diputados (cámara baja) y por el Senado (cámara alta). El Congreso se compone de un mínimo de 300 diputados y un máximo de 400, elegidos por sufragio universal, libre, igual, directo y secreto, por un periodo de cuatro años. En la actualidad el Congreso está compuesto por 350 diputados. (Del Valle y Gómez de Agüero, 2005, p.9-10)

La principal función del Congreso de los Diputados es la de elaborar los proyectos y proposiciones de ley que posteriormente pasan al Senado para que este estudie, modifique o vete dichos proyectos.

Tal y como se recoge en la Constitución española (1978): “El castellano es la lengua española oficial del Estado. Todos los españoles tienen el deber de conocerla y el derecho a usarla. Las demás lenguas españolas serán también oficiales en las respectivas Comunidades Autónomas de acuerdo con sus Estatutos [...]”. Sin embargo, el castellano es la única lengua utilizada en el Congreso de los Diputados, a diferencia de lo que sucede en el Parlamento Europeo donde se pueden utilizar las 24 lenguas oficiales.

Metodología

En el presente apartado metodológico y con arreglo a los usos habituales de la LC, diferenciaremos varias fases: metodología de compilación, metodología de preparación del corpus y metodología de análisis.

La metodología de compilación consiste en la recopilación del corpus. Puesto que nuestro trabajo consiste en el análisis de un corpus ya existente, ECPC (Corpus Comparables y Paralelos de Discursos Parlamentarios Europeos), para el presente trabajo no tuvimos que realizar ningún proceso de compilación. Aprovecharemos, no obstante, este espacio para presentar al equipo responsable del corpus que analizo aquí y para describir rasgos básicos de dicho corpus.

El grupo ECPC, cuya coordinadora es María Calzada (Universitat Jaume I), nace con prestigiosos investigadores de diferentes universidades europeas: Dublin City University (Irlanda), Durham University (Reino Unido), Trinity College Dublin (Irlanda) o Università di Bologna, entre otros. Entre ellos cabe destacar: la Catedrática Mona Baker, la Dra. Silvia Bernardini, la Dra. Dorothy Kenny y el Dr. Saturnino Luz, entre otros. Es, por tanto, un equipo multidisciplinar y plurinacional que se encarga de analizar discursos pronunciados en las sesiones plenarias de varios parlamentos europeos: el Parlamento Europeo (PE), el Congreso de los Diputados español (CD) y la Cámara de los Comunes británica (HC).

Basándonos en la clasificación teórica anterior, este corpus sería específico puesto que recoge discursos de contextos especializados (Congreso de los Diputados y Parlamento Europeo); grande, ya que se compone de numerosos discursos pronunciados a lo largo de bastantes años; etiquetado, puesto que con la aportación que he realizado se ha dotado al corpus de un etiquetado morfológico; bilingües, puesto que sus discursos son en inglés y español; y por último, comparable ya que en los textos aparecen rasgos como por ejemplo el tema, el tipo de texto, la función comunicativa o el grado de especialización.

El corpus ECPC se compone de discursos pronunciados en el Parlamento Europeo (en inglés, español original y traducido), el Congreso de los Diputados (español) y la House of Commons (inglés). Los discursos están en formato .xml y por ello, en cada uno de los archivos aparecen diferentes partes organizadas con diferentes etiquetas, entre las principales encontramos: *header*, *body*, *back*, *omit* y *post*. La etiqueta *<header>* contiene información metatextual acerca del discurso: fecha en la que se celebra, lugar, el título del discurso, etc. En la etiqueta *<body>* aparecen las intervenciones de cada interlocutor con información adicional: idioma de origen del discurso, nombre del ponente, etc. La etiqueta *<back>* está relacionada con temas de copyright y actualizaciones. La etiqueta *<omit>* nos proporciona datos concretos de la sesión, por ejemplo: la hora determinada en la que se abre la sesión, si se aplaza la sesión, etc. Por último, la etiqueta *<post>* nos especifica el cargo de cada una de las personas que interviene en los discursos, por ejemplo, presidente.

Cada uno de los archivos es un día en cada una de las instituciones. Cada corpus tiene extensiones diferentes. En el caso del corpus del PE se compone de 300 textos y 17.094.820 palabras mientras que el corpus del CD se compone de 499 textos y 15.772.512 palabras.

En concreto nuestro trabajo se centra en el análisis (el español original y traducido) de los discursos del PE y del español de los discursos del CD, comprendidos entre 2004 y 2009. Es importante señalar que en el PE, a partir del año 2004 (y hasta 2011, cuando ya no se publica la traducción oficial de los debates), todos los discursos se traducían primero al inglés y de ahí al resto de lenguas comunitarias.

En la metodología de preparación, destacaremos la aportación que realizamos al corpus ECPC antes de empezar a analizarlo. Esta fue el etiquetado morfológico, también conocido por sus siglas en inglés POS *tagging* (*part of speech tagging*), que se refiere al proceso de asignar a cada una de las palabras del corpus su categoría gramatical. Los programas o herramientas de etiquetado morfológico incorporan un *tagset* con abreviaturas para cada categoría. Entre las herramientas de etiquetado más corrientes o utilizadas encontramos: *Freeling*, *TagAnt* y *Treetagger*.

Freeling es un sistema desarrollado en la Universitat Politècnica de Catalunya bajo la dirección de Lluís Padró. El programa incluye los idiomas de español, catalán, gallego, inglés e italiano, entre otros.

TagAnt es una útil interfaz de etiquetado morfológico de la que es responsable Laurence Anthony y que está basada en el programa *Treetagger*.

Treetagger es una herramienta diseñada por el alemán Helmut Schmid en el Instituto de Lingüística Computacional de la Universidad de Stuttgart. Esta herramienta se ha usado con éxito para etiquetar en alemán, inglés, francés, español, ruso, entre muchos otros idiomas.

Tras explorar los tres programas decidimos utilizar *Treetagger* ya que, por un lado, es el más rápido y completo y, por otro lado, dispone de *tagsets* multilingües comparables, con etiquetas/abreviaturas para distintos idiomas.

Por último, para la metodología de análisis, que es la más extensa, la herramienta que utilizamos primordialmente ha sido *WordSmith Tools 6.0*. Este programa fue creado por Mike Scott en 1996. “WordSmith Tools is an integrated suite of programs for looking at how words behave in texts. You will be able to use the tools to find out how words are used in your own texts, or those of others” (Scott, 2015, p.17).

Entre los procesos para el análisis de corpus que comentaremos a continuación se encuentran: la valoración estadística de los corpus (con especial referencia a su grado de variación léxica), la creación de listas de frecuencia (*wordlists*), la detección automática de palabras clave (*keywords*) y la generación de concordancias (*concordances*). Además de estos tres procesos, analizaremos también el etiquetado morfológico que realizamos como aportación al corpus. En el análisis fusionaré los procedimientos de palabras clave y concordancias cuando sea necesario, por ejemplo, para esclarecer información procedente de las palabras clave.

A lo largo de todos los resultados mostraremos solamente algunas figuras y tablas ilustrativas, por motivos de espacio, el resto de figuras completas (listas de frecuencia, palabras clave, etc.) se podrán ver en el CD adjunto.

Variación léxica

Entre las estadísticas que WordSmith Tools proporciona al usuario cabe destacar el *standardised TTR (types tokens ratio)* que se explica así:

If a text is 1,000 words long, it is said to have 1,000 "tokens". But a lot of these words will be repeated, and there may be only say 400 different words in the text. "Types", therefore, are the different words. But this type/token ratio (TTR) varies very widely in accordance with the length of the text – or corpus of texts – which is being studied. (Scott, 2015, p.299)

En definitiva, este valor nos muestra la cantidad de palabras diferentes que tiene cada corpus y, por tanto, la riqueza o variación léxica de este, ya que cuanta más variedad, más rico es el lenguaje. El valor de variedad léxica que veremos a continuación en los resultados, es un valor que se obtiene en función del número total de palabras de cada corpus, mediante un cálculo que realiza el programa.

Listas de frecuencia

Automatic frequency list generation can quickly produce a complete list of all the items in a corpus, ranging from the most ubiquitous ones, the frequency of which may run into millions in the largest corpora, to those more unusual items which occur just once in a particular. (Evison, 2010, p.122)

Tras el examen de valores estadísticos, realizamos una lista de frecuencia para cada uno de los corpus de nuestro interés: una del español del CD y dos correspondientes al PE (de español original, la primera y de español traducido, la segunda).

La elaboración de las lista de frecuencias requirió no solo ciertos conocimientos de .xml (a partir del que se realiza la selección de los subcorpus de debates originales y traducidos) sino además dominio “(cuasi) experto” de WordSmith Tools. La idea a grandes rasgos consistía en seleccionar las partes concretas de cada corpus que nos interesaban y descartar las que carecían de interés. Para explicar los detalles, conviene aclarar que procedimos del modo siguiente. Primero, abrimos un texto de cada corpus con el editor *Notepad* (para visualizar el formato .xml) y así saber seleccionar exactamente qué información nos interesaba que apareciera en nuestra lista de frecuencia y qué información no. En los ajustes avanzados de *WordSmith Tools*, y en el apartado *only part of file*, nos aparecen dos columnas: *sections to cut* y *sections to keep*. La primera columna nos sirve para decir qué información no queremos que aparezca en nuestra lista de frecuencia porque no nos interesa. La segunda columna, *sections to keep*, nos permite incluir qué información sí queremos que aparezca en nuestra lista de frecuencias. Dado que lo que interesa estudiar son las intervenciones de los diputados en todos los corpus habrá algunas partes que no nos darán información relevante y que, por tanto, omitiremos. Estas partes que no presentan información relevante para el estudio serán siempre estas cuatro etiquetas: *<header>*, *<back>*, *<omit>* y *<post>*.

Además de estas especificaciones generales a todos los corpus de ECPC, cada corpus tendrá otras restricciones particulares que analizaremos a continuación.

En el caso del CD, aparte de eliminar las etiquetas que comentamos anteriormente, solo pondremos en la sección de *sections to keep* la etiqueta *<speech>* para marcar que lo que queremos seleccionar del corpus es lo que incluye esa etiqueta, que básicamente son las intervenciones de los diputados.

Para crear la lista de frecuencia del Parlamento en español original, la restricción más importante fue la del idioma. La etiqueta *<language>* nos marca el idioma original del discurso. Dado que el idioma que queríamos analizar era el español original únicamente, en el apartado *sections to keep* introdujimos la etiqueta *<language="ES">*, que indica que la intervención que le sigue es español original.

En el caso del Parlamento en español traducido, teníamos que marcar lo contrario que el anterior. Lo que nos interesaba era cualquier idioma original que no fuera español, es decir, una traducción. Para ello, la etiqueta anterior (*<language="ES">*) la introdujimos en *sections to cut*, para indicar que no utilizara palabras procedentes de las intervenciones en español original.

Palabras clave

Las palabras clave representan los vocablos más idiosincrásicos de un corpus respecto de otro, generalmente más grande, que denominamos corpus de referencia. “A key-word list includes items that are either significantly frequent (positive key words) or infrequent (negative key words)” (Evison, 2010, p.127).

El segundo paso consistió en realizar las listas de palabras clave. Tal y como se explicó antes, para la creación de listas de palabras clave tuvimos que comparar dos corpus, concretamente dos listas de frecuencia. Para ello, hicimos uso, principalmente, de las listas de frecuencia del apartado anterior.

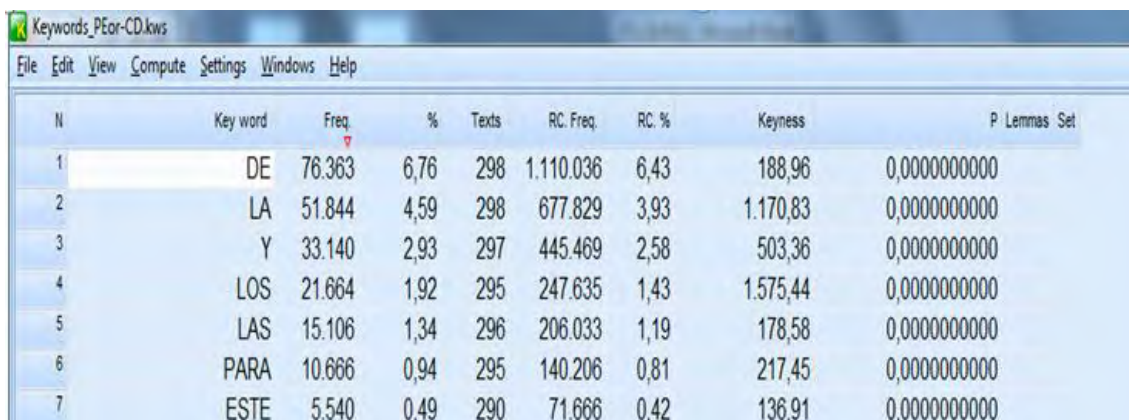
En primer lugar realizamos una comparación entre la lista de frecuencia del PE original y la lista de frecuencia del CD y viceversa; y luego, tanto la del PE y el CD con otro corpus más extenso, el de la RAE.

Elegimos comparar tanto el CD como el PE con la RAE, ya que ésta presenta las palabras más utilizadas en un contexto normal y no especializado, como es el caso de las otras dos instituciones donde nos encontramos en contextos más políticos.

En las listas de palabras claves, se necesita realizar un proceso de normalización como explica Evison:

In order to compare frequency counts across corpora of different sizes, a process of normalization is required. This process involves extrapolating raw frequencies from the different-sized corpora which are being compared so that they can be expressed by a common factor such as a thousand or a million words. (Evinson, 2010, p.126)

Mediante WordSmith 6.0 se automatiza el proceso de normalización del que habla Evinson y ello da lugar a una lista de palabras clave a partir de cálculos basados en los porcentajes de aparición de los vocablos tanto en el corpus examinado como en el de referencia. Dicho de otro modo, la columna llamada *Freq.* (véase Figura 1) indica el número de veces que esa palabra aparece en el corpus y el símbolo % se refiere a la proporción en relación al total de palabras. Por tanto, es la columna del porcentaje y no la de frecuencia en la que deberemos basarnos para comparar un corpus propio con un corpus de control.



N	Key word	Freq.	%	Texts	RC. Freq.	RC. %	Keyness	P. Lemmas Set
1	DE	76.363	6,76	298	1.110.036	6,43	188,96	0,0000000000
2	LA	51.844	4,59	298	677.829	3,93	1.170,83	0,0000000000
3	Y	33.140	2,93	297	445.469	2,58	503,36	0,0000000000
4	LOS	21.664	1,92	295	247.635	1,43	1.575,44	0,0000000000
5	LAS	15.106	1,34	296	206.033	1,19	178,58	0,0000000000
6	PARA	10.666	0,94	295	140.206	0,81	217,45	0,0000000000
7	ESTE	5.540	0,49	290	71.666	0,42	136,91	0,0000000000

Figura 1. Lista de palabras clave

Concordancias

A concordance is a collection of the occurrences of a word-form, each in its own textual environment. In its simplest form it is an index. Each word-form is indexed and a reference is given to the place of occurrence in a text. (Sinclair, 1991, p.32)

En WordSmith Tools, entre las opciones más interesantes asociadas a las concordancias aparecen: *concordances*, *collocates* y *clusters*. En primer lugar, la opción *concordance* nos muestra el vocablo en cotexto (el contexto inmediato donde se muestran las palabras de antes y después) y en contexto (más amplio). Los *collocates* son “listados de palabras que aparecen alrededor de la palabra base, en posiciones determinadas” (Vargas, 2012, p.73). Los *clusters* nos muestran los grupos de (tres)

palabras más utilizados en los que de manera implícita o explícita está incluida la palabra que hemos querido estudiar.

Una vez realizamos las listas de palabras clave, es cuando nos surgieron las posibles ideas para analizar palabras concretas. En este caso, analizaremos las palabras: *Derechos, Señor, Señora, Gracias y Usted*.

Etiquetado morfológico

Por último, analizamos el corpus desde el punto de vista gramatical gracias al etiquetado morfológico (POS), identificando la cantidad de adjetivos, nombres, conjunciones coordinadas, verbos, pasiva refleja y verbos modales que tienen cada uno de los corpus. El procedimiento específico de estos análisis fue el siguiente. Cargamos en *WordSmith Tools* los diferentes textos que tenían el etiquetado morfológico (y que constituían corpus independientes de sus equivalentes en bruto, sin POS). A continuación, en la opción de *Concord* del programa introducimos la abreviatura de cada categoría para así obtener la lista de todos los contextos donde aparece y, de este modo, poder saber el número de veces que aparecía ese tipo de categoría gramatical. Sabíamos la abreviatura de cada categoría gramatical gracias al *tagset* del programa de etiquetado *Treetagger*.

Resultados

En este apartado mostraremos los resultados de cada uno de los procedimientos de análisis que se han descrito arriba.

Variación léxica

Entre los datos estadísticos de nuestros corpus, estos son los valores del Standardised TTR (y por consiguiente del grado de variación léxica de cada corpus):

Tabla 1. Variación léxica de los corpus

Corpus	Variación léxica
Congreso de los Diputados	39,69
Parlamento Europeo (original)	42,13
Parlamento Europeo (traducido)	42,61

Aunque los resultados del Parlamento tanto original como traducido son muy próximos, la lengua traducida es la que presenta mayor variación léxica. Tal vez esto se podría explicar, como veremos más adelante, con el hecho de que, en muchas ocasiones, la traducción modera el elevado uso de ciertas categorías gramaticales en originales de contextos multilingües, como la abundancia de conjunciones coordinadas. Podría ocurrir que, con tal fin de moderación, se recurra a otro tipo de palabras o a fraseología más idiomática del español y ello redunde en una variación léxica mayor.

Listas de frecuencia

Parlamento europeo. Aquí se muestra una parte de las listas de frecuencia del PE (tanto el español original como el traducido). Para ver las listas de frecuencia completas véase el apartado 1.2 Parlamento Europeo (original) y 1.3 Parlamento Europeo (traducido) del CD adjunto.

N	Word	Freq.	%	Texts	% Lemmas	S
72	ACUERDO	1.427	0,13	244	81,33	
73	TIENE	1.396	0,12	252	84,00	
74	LUGAR	1.367	0,12	248	82,67	
75	NA	1.347	0,12	70	23,33	
76	HACER	1.319	0,12	247	82,33	
77	CONTRA	1.300	0,12	230	76,67	
78	DEBE	1.298	0,11	242	80,67	
79	DESARROLLO	1.280	0,11	218	72,67	
80	TENEMOS	1.253	0,11	244	81,33	
81	IMPORTANTE	1.251	0,11	251	83,67	
82	QUIERO	1.216	0,11	258	86,00	
83	ESTOS	1.205	0,11	249	83,00	
84	PUEDE	1.192	0,11	253	84,33	
85	ESTAMOS	1.186	0,11	246	82,00	
86	HOY	1.174	0,10	239	79,67	
87	YO	1.171	0,10	243	81,00	
88	ELLO	1.143	0,10	259	86,33	
89	ESE	1.132	0,10	246	82,00	
90	HECHO	1.128	0,10	246	82,00	
91	SITUACIÓN	1.121	0,10	226	75,33	
92	PROPUESTA	1.105	0,10	222	74,00	
93	CIUDADANOS	1.097	0,10	212	70,67	
94	SIDO	1.092	0,10	257	85,67	
95	ESO	1.082	0,10	241	80,33	
96	SEGURIDAD	1.071	0,09	220	73,33	
97	AÑO	1.054	0,09	204	68,00	
98	ASÍ	1.046	0,09	247	82,33	
99	SINO	1.046	0,09	246	82,00	
100	AÑOS	1.045	0,09	220	73,33	

Figura 2. Lista de frecuencia del Parlamento Europeo (original).

N	Word	Freq.	%	Texts	%	Lemmas	Set
67	SEÑORA	20.783	0,13	294	98,00		
68	HEMOS	20.457	0,13	298	99,33		
69	ACUERDO	20.167	0,13	298	99,33		
70	PUEDE	19.239	0,12	298	99,33		
71	PARTE	19.133	0,12	298	99,33		
72	HECHO	19.057	0,12	298	99,33		
73	DESARROLLO	18.190	0,11	296	98,67		
74	DEBE	18.104	0,11	298	99,33		
75	HAY	17.737	0,11	298	99,33		
76	CIUDADANOS	17.438	0,11	298	99,33		
77	DEBATE	17.298	0,11	298	99,33		
78	HACER	17.170	0,11	298	99,33		
79	CUESTIÓN	17.102	0,11	298	99,33		
80	AHORA	17.065	0,11	298	99,33		
81	TIENE	17.054	0,11	298	99,33		
82	TRABAJO	16.915	0,11	298	99,33		
83	CREO	16.842	0,11	298	99,33		
84	DEBEMOS	16.606	0,10	298	99,33		
85	SEGURIDAD	16.548	0,10	298	99,33		
86	ASÍ	16.483	0,10	298	99,33		
87	QUIERO	16.441	0,10	297	99,00		
88	NUESTRA	16.388	0,10	298	99,33		
89	TENEMOS	16.238	0,10	298	99,33		
90	CUANDO	16.207	0,10	298	99,33		
91	RESPECTO	16.048	0,10	298	99,33		
92	CONTRA	15.998	0,10	298	99,33		
93	FORMA	15.881	0,10	297	99,00		
94	PORQUE	15.755	0,10	298	99,33		
95	SOLO	15.467	0,10	290	96,67		
frequency	alphabetical	statistics	filenames	notes			

Figura 3. Lista de frecuencia del Parlamento Europeo (traducido).

Entre las listas de frecuencia del PE (tanto español original como traducido) observamos que no hay diferencias muy marcadas o significativas. Las palabras son prácticamente iguales, solo varía la frecuencia de aparición. Estas listas nos proporcionan información de carácter puramente conceptual. Las primeras palabras que aparecen son funcionales: preposiciones, artículos, etc. No obstante, entre las palabras más frecuentes identificamos algunas propias del contexto institucional en el que nos encontramos: *comisión, europea, unión, presidente*, etc. A continuación encontramos otras que nos aportan información sobre los asuntos de los que se habla más en este organismo, por ejemplo: *derechos, desarrollo, ciudadanos, trabajo, humanos, cooperación, protección, compromiso, pobreza*, etc. Por estas palabras podemos deducir que se tratarán temas sociales como los derechos humanos, temas relacionados con los ciudadanos, con la cooperación internacional o la pobreza.

Congreso de los Diputados. A continuación se muestra una parte de la lista de frecuencias del Congreso de los Diputados. Para ver la lista completa, véase el apartado 1.1 Congreso de los Diputados del CD adjunto.

N	Word	Freq.	%	Texts	% Lemmas	Set
66	POLÍTICA	24.651	0,14	443	98,66	
67	POPULAR	24.421	0,14	445	99,11	
68	SEÑORÍAS	24.381	0,14	446	99,33	
69	ESO	23.571	0,14	443	98,66	
70	CUANDO	23.259	0,13	444	98,89	
71	SIN	22.893	0,13	446	99,33	
72	ESA	22.535	0,13	444	98,89	
73	HACER	22.069	0,13	444	98,89	
74	NI	21.920	0,13	442	98,44	
75	PARTIDO	21.746	0,13	443	98,66	
76	VA	21.707	0,13	445	99,11	
77	ESPAÑA	21.674	0,13	442	98,44	
78	ESE	21.447	0,12	445	99,11	
79	ESTAMOS	21.006	0,12	444	98,89	
80	QUÉ	21.006	0,12	440	98,00	
81	ENTRE	20.080	0,12	444	98,89	
82	PARTE	19.909	0,12	444	98,89	
83	CÁMARA	19.731	0,11	449	100,00	
84	MINISTRO	19.712	0,11	436	97,10	
85	AÑOS	19.702	0,11	443	98,66	
86	CIENTO	19.395	0,11	433	96,44	
87	HECHO	19.184	0,11	446	99,33	
88	NOSOTROS	19.084	0,11	444	98,89	
89	ESTOS	18.698	0,11	443	98,66	
90	PUEDE	18.193	0,11	444	98,89	
91	SOCIALISTA	17.842	0,10	444	98,89	
92	DECIR	17.830	0,10	442	98,44	
93	ESTO	17.597	0,10	442	98,44	
94	DOS	17.519	0,10	447	99,55	

Figura 4. Lista de frecuencia Congreso de los Diputados

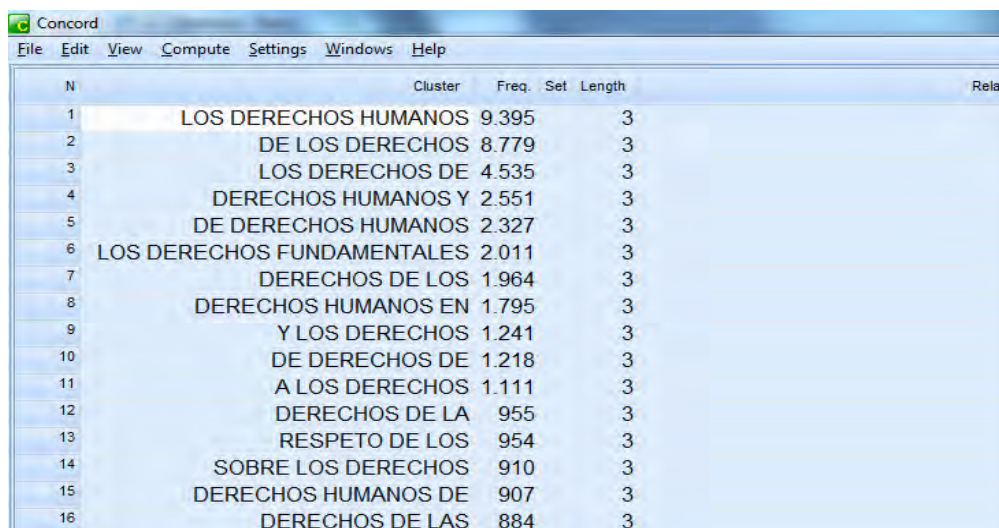
En la lista observamos entre las palabras con mayor frecuencia una serie de palabras propias del contexto político nacional, como por ejemplo: *Señor, gobierno, ley, presidente, popular, partido, cámara, español, ministro, comunidades, Cataluña* etc. A continuación, conforme van avanzando las entradas obtenemos palabras como: *millones, euros, derechos, reforma, presupuestos, economía, sociedad, empleo*, etc. Estas palabras nos hacen deducir que algunos de los temas más frecuentes en el Congreso de los Diputados son de índole socioeconómica.

Palabras clave

En este apartado presentamos las diferentes listas de palabras clave y, cuando resulta oportuno, añadimos también el procedimiento de concordancias para clarificar información procedente de las palabras clave.

La primera lista de palabras clave la generamos entre el PE (español original) y el CD. La lista nos vuelve a confirmar lo que nos había mostrado la lista de frecuencia, que las palabras más representativas del PE están más relacionadas con temas sociales y que, por supuesto, tiene un índice más alto de aparición que en el Congreso. Algunos ejemplos de estas palabras son: *derechos, desarrollo, trabajo, humanos, apoyo*, etc.

Esto lo podemos ver con el ejemplo de la palabra *derechos*. Si realizamos las concordancias de esta palabra primero en el PE y luego en el CD, obtenemos resultados muy diferentes. En el PE, tanto la opción de los *clusters* como la de *collocates*, nos muestra que, en la gran mayoría de los casos, la palabra *derechos* va acompañada de *humanos* o *fundamentales* y seguido de una referencia geográfica. A continuación, se muestran los *clusters* y los contextos donde aparece dicha palabra. Para ver las listas completas, véase el apartado 3.1 Palabra derechos PE del CD adjunto.



N	Cluster	Freq.	Set	Length	Relate
1	LOS DERECHOS HUMANOS	9.395		3	
2	DE LOS DERECHOS	8.779		3	
3	LOS DERECHOS DE	4.535		3	
4	DERECHOS HUMANOS Y	2.551		3	
5	DE DERECHOS HUMANOS	2.327		3	
6	LOS DERECHOS FUNDAMENTALES	2.011		3	
7	DERECHOS DE LOS	1.964		3	
8	DERECHOS HUMANOS EN	1.795		3	
9	Y LOS DERECHOS	1.241		3	
10	DE DERECHOS DE	1.218		3	
11	A LOS DERECHOS	1.111		3	
12	DERECHOS DE LA	955		3	
13	RESPETO DE LOS	954		3	
14	SOBRE LOS DERECHOS	910		3	
15	DERECHOS HUMANOS DE	907		3	
16	DERECHOS DE LAS	884		3	

Figura 5. *Clusters* de la palabra *derechos* en el Parlamento Europeo (original)



N	Concordance	Set	Tag	Word #	Sen	Sen	Para	Para H
1	la sesión anterior: véase el Acta 10. Derechos humanos en el mundo 2005			189	19	11	0	0%
2	la sesión anterior: véase el Acta 10. Derechos humanos en el mundo 2005			39.30	1.2	2%	0	37
3	(debate) 11.1. Chad oriental 11.2. Derechos de las mujeres en Arabia			288	28	22	0	1%
4	de los debates. (1)Cf. Acta. 11.2. Derechos de las mujeres en Arabia			32.02	1.0	4%	0	82
5	en Transdniéster (Moldova) 11.3. Derechos humanos en Vietnam 12.			281	31	33	0	1%
6	el problema de Transdniéster. 11.3. Derechos humanos en Vietnam			44.05	1.5	4%	0	92
7	común (RC-B6-0527/2007) 12.2. Derechos de las mujeres en Arabia			38.78	1.3	13	0	99
8	12.1. Chad oriental (votación) 12.2. Derechos de las mujeres en Arabia			311	32	20	0	1%
9	(versión refundida) (debate) 13. Derechos de los pasajeros que viajan			808	55	8%	0	1%
10	lugar mañana a las 12.00 horas. 13. Derechos de los pasajeros que viajan			72.50	2.3	1%	0	60
11	(Moldova) (votación) 13.3. Derechos humanos en Vietnam			320	36	29	0	1%
12	aprueba las enmiendas orales 13.3. Derechos humanos en Vietnam			47.73	1.6	6%	0	99
13	una ofensiva propagandística. 14.2. Derechos humanos en Etiopía			30.31	1.0	5%	0	81
14	de Derecho(debate) 14.1. Cuba 14.2. Derechos humanos en Etiopía 14.3.			272	27	33	0	1%
15	Televisión" en Venezuela 14.2. Derechos humanos en Siria 14.3.			216	25	33	0	1%
16	los debates. (1) Véase el Acta. 14.2. Derechos humanos en Siria			27.60	1.0	4%	0	76

Figura 6. Contextos de la palabra *derechos* en el Parlamento Europeo (original)

Sin embargo, en el CD aunque la palabra también aparece acompañada por *fundamentales* en muchas ocasiones, su uso no es tan frecuente ni está tan centrado en los derechos humanos en el mundo. Además de ir acompañada de ésta, encontramos más variedad que va marcando ese campo político y económico diferenciado del PE,

como por ejemplo: *derechos económicos, derechos políticos o derechos constitucionales*. Para ver las listas completas, véase el apartado 3.2 del CD adjunto.

Seguidamente, hemos realizado la lista de palabras clave entre el CD y el PE. Aquí se muestra una parte de esta lista:

N	Key word	Freq.	%	Texts	RC. Freq.	RC. %	Keyness	P Lemmas	Set
1	QUE	750.420	4,35	448	43.87	3,89	565,93	0,000000	
2	EL	480.655	2,78	449	30.11	2,67	55,84	0,000000	
3	A	377.910	2,19	449	23.48	2,08	61,21	0,000000	
4	DEL	237.063	1,37	449	13.04	1,16	395,86	0,000000	
5	NO	208.582	1,21	449	10.18	0,90	920,09	0,000000	
6	POR	203.622	1,18	449	11.54	1,02	237,63	0,000000	
7	SE	201.595	1,17	449	10.51	0,93	556,36	0,000000	
8	#	164.875	0,96	449	9.257	0,82	217,01	0,000000	
9	LO	133.966	0,78	448	5.733	0,51	1.139,72	0,000000	
10	HA	92.194	0,53	448	5.223	0,46	107,71	0,000000	
11	SEÑOR	90.935	0,53	447	4.863	0,43	200,72	0,000000	
12	ESTA	76.453	0,44	446	3.515	0,31	468,67	0,000000	
13	GOBIERNO	68.397	0,40	446	684	0,06	4.824,42	0,000000	
14	GRUPO	57.886	0,34	445	765	0,07	3.454,88	0,000000	
15	LEY	55.181	0,32	443	181	0,02	5.581,53	0,000000	
16	GRACIAS	51.516	0,30	446	452	0,04	3.870,40	0,000000	
17	PERO	50.664	0,29	445	2.870	0,25	59,12	0,000000	
18	PORQUE	49.600	0,29	446	1.875	0,17	649,22	0,000000	
19	O	48.437	0,28	447	2.342	0,21	224,94	0,000000	
20	NOS	45.620	0,26	444	2.181	0,19	226,33	0,000000	
21	USTED	39.258	0,23	441	584	0,05	2.146,98	0,000000	
22	SI	37.722	0,22	448	1.895	0,17	136,91	0,000000	
23	TIENE	36.383	0,21	446	1.396	0,12	454,62	0,000000	
24	LE	35.137	0,20	445	761	0,07	1.331,34	0,000000	
25	ESTADO	33.819	0,20	443	853	0,08	1.043,01	0,000000	
26	MUCHAS	33.533	0,19	446	571	0,05	1.636,32	0,000000	
27	YA	32.179	0,19	446	1.583	0,14	134,22	0,000000	
28	SEÑORA	31.979	0,19	442	1.582	0,14	128,65	0,000000	
29	DESDE	30.711	0,18	445	1.594	0,14	87,39	0,000000	

Figura 7. Lista de palabras claves Congreso de los Diputados y Parlamento Europeo.

En este caso lo que más llama la atención son las palabras propias del contexto español y, por tanto, más características del CD frente al PE: *gobierno, ley, popular, España, ministro, socialista, comunidad, autónomas, Cataluña, etc.*

Por otra parte, a simple vista llaman la atención palabras que se enmarcan dentro de un registro formal y de cortesía como: *señor, gracias, señora y usted*. Esto quiere decir que aparecen en una proporción significativa y mayor que en el PE. Frente a posibles ideas preconcebidas acerca de que la cortesía europea es superior a la española, con este análisis la concepción cambia totalmente. Hemos consultado las concordancias de las palabras *gracias* y *señor* para ver qué diferencias había en cada contexto. En cuanto a la palabra *gracias* la opción de *clusters* y la de concordancias nos muestran cómo en el CD se utiliza de forma más directa mediante el acompañamiento del vocativo, como muestra la colocación más frecuente: *Gracias, Señor Presidente*. Sin embargo, en el PE se usa de una forma más indirecta, acompañada de perífrasis de agradecimiento, como por ejemplo: *Quiero dar las gracias, doy las gracias, gracias a, etc.* Para consultar estas listas, véanse los apartados 3.3 y 3.4 del CD adjunto.

En cuanto a la palabra *señor*, cabe destacar que en el CD se utiliza en la mayoría de las veces seguido de *gracias*: *Gracias Señor Presidente*, *Muchas gracias Señor*, etc. En el PE, la colocación más frecuente de la palabra es: *Señor Presidente*, usado como vocativo. Por tanto, podemos decir que el CD agradece mucho más a lo largo de sus discursos y lo hace de forma directa, al contrario que el PE que utiliza mucho menos el agradecimiento con expresión indirecta.

Por último, hemos realizado las últimas listas de palabras clave. Esta vez hemos utilizado la lista de frecuencia de la RAE, es decir, del lenguaje común. La primera lista la hemos realizado entre el PE original y la RAE. La lista nos muestra las palabras más características del PE frente al lenguaje común. Observamos que son palabras típicas del discurso parlamentario y que algunas ni siquiera se utilizan en el lenguaje común y, por tanto, su frecuencia en la lista de la RAE es de cero. Algunas de estas palabras son: *votado* o *comisaria*.

En el caso de la lista de palabras clave entre la RAE y el PE se muestran las palabras más idiosincrásicas del lenguaje común. No tiene nada que ver con las palabras de la lista anterior y observamos que aparecen bastantes nombres propios típicos como: *Juan*, *José* o *María*. Por otro lado, las palabras que aparecen son palabras cotidianas y sencillas utilizadas en contextos diarios, como por ejemplo: *hombre*, *casa*, *ciudad*, *madre*, *cuerpo*, etc. La frecuencia de las palabras *música* o *amor* nos muestran que son temas característicos del lenguaje común, pero que se alejan totalmente de los temas del PE. También destacaremos que hay algunas palabras con frecuencia cero en el PE, como por ejemplo: *pesetas*, *pintura*, *copa* o *habitación*.

A continuación hemos cruzado la lista de frecuencia del CD y de la RAE. En este caso volvemos a encontrarnos con las palabras propias del CD frente al lenguaje común: *señor*, *gracias*, *ley*, *señorías*, *gobierno*, etc. Otra vez más encontramos palabras con frecuencia cero en la RAE, y por tanto, que no se utilizan nada en el lenguaje común, como por ejemplo; *señoría*, *emitidos* o *abstenciones*.

En el caso de la lista entre la RAE y el CD, encontramos que, en relación a las palabras propias del lenguaje común, ocurre prácticamente lo mismo que con las listas de la RAE y el PE. Las palabras son prácticamente las mismas: *hombre*, *casa*, *noche*, etc.

Cabría destacar que todas las palabras del lenguaje común aparecen siempre en el CD, aunque en muchas ocasiones la frecuencia es mínima. Esto no ocurre así en el PE, ya que en ocasiones estas palabras del lenguaje común no aparecen ni una sola vez en dicho organismo.

Etiquetado morfológico (*POS*)

A continuación se muestra la gráfica con los resultados de las siguientes cinco categorías gramaticales de los corpus: adjetivos, sustantivos, conjunciones, verbos, pasiva refleja con *se* y verbos modales.

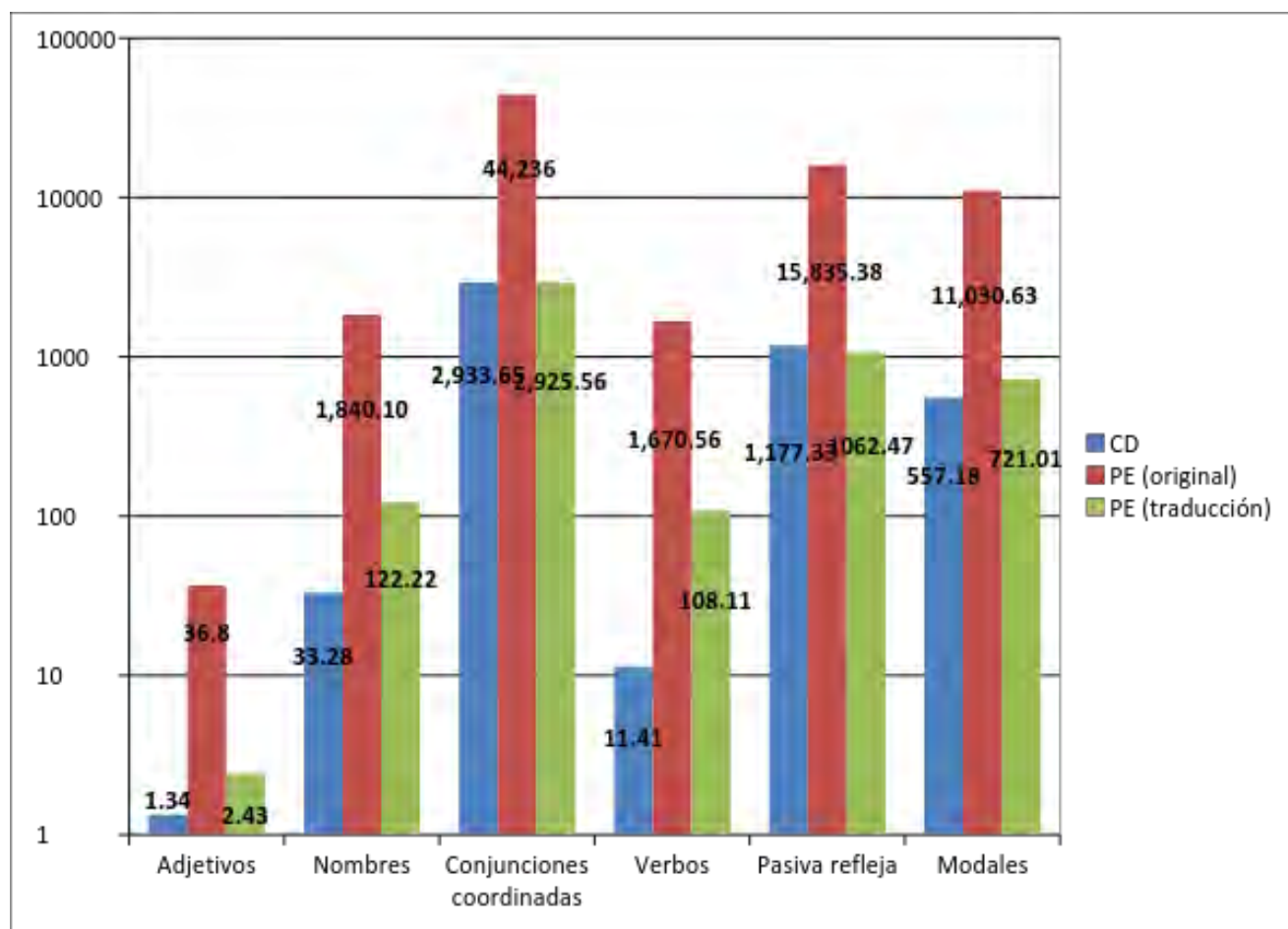


Figura 8. Frecuencia de las categorías gramaticales en cada organismo.

En un primer momento, observamos que el PE (original) contiene el número más alto de todas las categorías gramaticales. A continuación, salvo en el caso de las

conjunciones coordinadas y de la pasiva refleja con *se*, la traducción se posiciona en segundo lugar de frecuencia.

De forma general, se observa que la traducción modera y reduce los usos tan elevados de todas las categorías gramaticales en el PE original. Además, la traducción se acerca más a los datos que se obtienen en el CD, es decir, al español original que no está en un contexto multilingüe influenciado por el inglés.

Los datos de los adjetivos, los verbos, los sustantivos y los verbos modales muestran que son muchos menos frecuentes en el CD. En el caso de los verbos modales, muy típicos del inglés, la traducción modera su utilización acercándose a las cifras del español del Congreso.

Las conjunciones coordinadas y la pasiva refleja presentan una excepción a esta tendencia que siguen el resto de categorías gramaticales. En cuanto a las conjunciones coordinadas, muy típicas del inglés, encontramos que la cifra en el CD y en el PE traducido es prácticamente la misma, aunque aun así, el CD presenta la frecuencia más elevada. Por una parte, el hecho de que la traducción tenga la frecuencia más baja de todas las categorías en cuanto a coordinación se puede explicar porque, en español, es más común utilizar la subordinación o signos de puntuación y por tanto, los traductores podrían utilizar este tipo de recursos para que quede un texto más idiomático. En este caso, dado que los traductores están al tanto de estas cuestiones, podrían detectar el uso abusivo de estas conjunciones coordinadas tan típicas del inglés y cambiarlas por otras formas como las que hemos explicado anteriormente. Aun así, esto es solo una hipótesis o conjetura que se tendría que confirmar con trabajos posteriores. En cuanto a la pasiva refleja, cierto es que el CD contiene la segunda frecuencia más alta de utilización, con una cifra cercana a la del Parlamento traducido. Esto puede deberse a que en estos dos contextos se utilicen alternativas a esta pasiva refleja como puede ser la forma activa. Una vez más, esto supone una mera especulación que debe ser comprobada con más estudios. Estos dos datos (tanto las conjunciones coordinadas como la pasiva refleja) cambian la tendencia de lo que se había visto hasta ahora con los datos anteriores.

Conclusiones

En este último apartado, por una parte destacaremos conclusiones específicas de los resultados del trabajo y por otra parte, conclusiones de tipo general acerca de para qué me ha servido este trabajo, aplicaciones futuras, etc.

En primer lugar, hemos constatado con este análisis que los corpus sirven para discernir (o detectar) perfectamente la terminología específica de un tema o situación concreta; en nuestro caso, el género de los debates parlamentarios. Gracias a las listas de frecuencia y las palabras clave podemos conocer dicha terminología rápidamente.

Desde el punto de vista conceptual, como hemos ido viendo en los resultados, concluimos que en el Parlamento Europeo se tratan temas más sociales mientras que en el Congreso de los Diputados se tratan temas de carácter más económico.

Al mismo tiempo, desde el punto de vista gramatical, observamos que la traducción modera en muchas ocasiones, el uso tan elevado que se emplea de las categorías gramaticales en discursos originales. Esto podría explicarse debido a la labor que realiza el traductor que, como buen conocedor de la lengua, adaptaría y reduciría usos gramaticales que en español no son tan idiomáticos o los cambiaría por otros recursos más propios, como es el caso de las conjunciones coordinadas, muy típicas del inglés pero no tanto del español (López Guix y Minett Wilkinson, 1997, p.89-91). La reducción de estas categorías gramaticales en el corpus traducido podría explicarse también con el hecho de que este presente una mayor variedad léxica; no podemos descartar la hipótesis de que los recursos gramaticales diferentes varíen y enriquezcan el léxico (y viceversa). No obstante, como ya hemos mencionado antes, todo esto supone una mera hipótesis que ha de corroborarse o refutarse en estudios futuros.

En el caso de que dichos estudios futuros corroboraran las primeras impresiones que se obtienen de los análisis de este trabajo de fin de grado, ello supondría la revisión de presuntos universales propuestos por varios teóricos como la simplificación y la explicitación. Si según Baker (1993): “Simplification is ‘the idea that translators subconsciously simplify the language or message or both’” (Baker, 1996, p.181), como los datos que se han examinado aquí nos indican que la traducción presenta la mayor variedad léxica, ello obligaría a revisar la simplificación.

En cuanto a la explicitación, Baker afirma que esta supone: “the tendency to spell things out in translation, including, in its simplest form, the practice of adding background information” (Baker, 1996, p.180). Por tanto, como en este caso los datos nos indican que la traducción presenta la menor frecuencia en cuanto a la coordinación, ello obliga a revisar la explicitación. No obstante, la explicitación abarca muchos más aspectos y no solo la coordinación (que es el aspecto que se ha examinado aquí), y esto habría de tenerse en cuenta a la hora de estudiar más en profundidad esta cuestión de la explicitación concreta.

En cuanto a las conclusiones generales del trabajo, cabe destacar que este nos ha servido para conocer una metodología con la que hasta este momento no habíamos trabajado en profundidad, la lingüística de corpus. Gracias a haber podido conocer qué herramientas son útiles para el análisis de corpus (como el caso de *TreeTagger* o *WordSmith tools*) y qué funciones presentan, hemos constatado que son de gran ayuda para el intérprete y el traductor y que, a partir de ahora, las tendremos en consideración.

Los corpus, y por supuesto sus herramientas de análisis, ofrecen tanto calidad como una mayor rapidez en el proceso de trasvase. Para el intérprete, herramientas como *WordSmith Tools* (y equivalentes libres como *AntConc*) resultan, en primer lugar, de gran utilidad para la documentación. Asimismo, sirven para proveer al profesional de la terminología y fraseología adecuada con un esfuerzo contenido. Para los traductores también resultan fundamentales ya que, como hemos indicado, pueden reforzar la calidad de las traducciones. Por un lado, igual que en el caso del intérprete, facilitan el uso de terminología específica. Por otro lado, ayudan a confirmar decisiones intuitivas que pudiéramos tener, proporcionan información sobre las colocaciones, respaldan el uso de nuevas expresiones, enriquecen el conocimiento sobre la lengua de llegada, etc. En definitiva, este tipo de herramientas permiten analizar una cantidad de corpus muy grande con rapidez y sistematicidad agilizan de forma considerable nuestro trabajo.

Bibliografía

- Del Valle, V y Gómez de Argüero, J.L. (2005) *Elementos de derecho*. Madrid: McGraw-Hill.
- ECPC. European Comparable and Parallel Corpora. Recuperado el 26 de enero de 2015, de <http://www.ecpc.uji.es/ES/home.php?language=es>
- Evison, J (2010). Analysing a corpus. What are the basics? En O’Keeffe, A & McCarthy, M. (Eds) *The Routledge handbook of corpus linguistics*. Londres: Routledge.
- Koskinen, K. (2000). Institutional Illusions. Translating in the EU Commission. *The Translator*, 6, (1), 49-65. Recuperado de http://www.academia.edu/7804246/_2000_Institutional_Illusions_Translating_in_the_EU_Commission
- López Guix, J.G y Minnet Wilkinson, J (1997) *Manual de traducción*. Barcelona: Gedisa.
- Parlamento Europeo. Recuperado el 28 de enero de 2015, de <http://www.europarl.es/>
- Sinclair, J.M (1991) *Corpus, Concordance and Collocation*. Oxford: Oxford University Press.
- Scott, M. (2015) *WordSmith Tool version 5*. Stroud: Lexical Analysis Software.
- Valdivieso, M. (Diciembre 2008). Política lingüística, norma lingüística y terminología en el plano supranacional. La Unión Europea. *Revista Electrónica de Estudios Filosóficos*. (16). Recuperado de

<https://www.um.es/tonosdigital/znum16/secciones/tritono-3-politica.htm> el 26 de enero de 2015.

Vargas, C. (2012). La tecnología de corpus en el contexto profesional y académico de la traducción y la terminología: panorama actual, recursos y perspectivas.

Tecnología, traducción y cultura, 67-99. Recuperado de

http://www.academia.edu/6252908/La_tecnolog%C3%ADa_de_corpus_en_el_contexto_profesional_y_acad%C3%A9mico_de_la_traducci%C3%B3n_y_la_terminolog%C3%ADa_panorama_actual_recursos_y_perspectivas el 5 de mayo de 2015.